

Evaluating the unseen five years on

ChatGPT, algorithms and machine
learning

Today's session

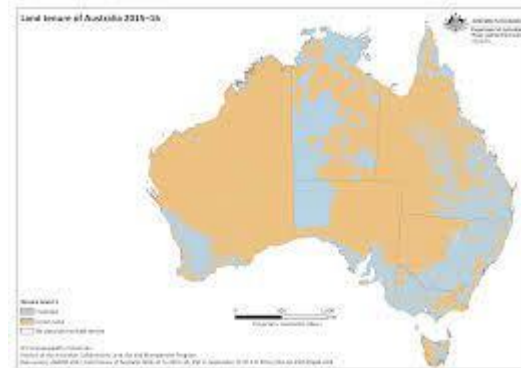
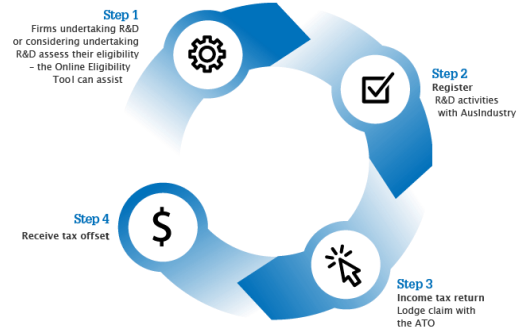
- 01 Why is this topic important?
- 02 Demystifying the unseen
- 03 Parameters for evaluating AI-enabled programs
- 04 Some tips to help you get started



What's my cred?

A decade in evaluation, assurance and governance

R&D Tax Incentive Program Cycle



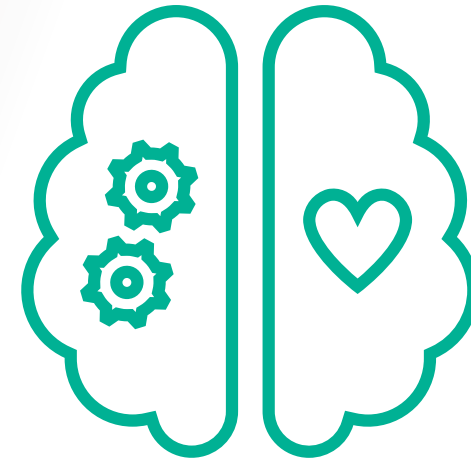
SPAT
SPEND • ANALYSIS • TOOL

Why is this topic important now?

- › ChatGPT released late last year, and upgraded again early this year
- › unfettered public access to extremely powerful generative AI LLM and ML
- › are we at the singularity?
- › locally, Robodebt
 - first known AI-enabled suicide
 - Neuralink closing the brain-computer divide
 - most development in private enterprises in the US and China
- › globally:

Before we go too far, an overview

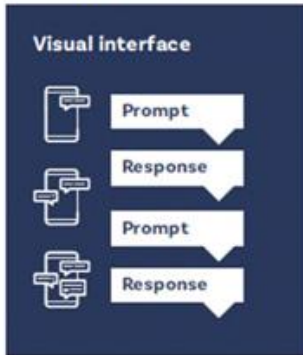
- › ADM (automated decision-making)
- › AGI (artificial general intelligence)
- › LLM (large language models)
- › MFM (multimodal foundation models)
- › ML (machine learning)



How ChatGPT works

Example LLM user experience
(based on ChatGPT-3)

What the user sees



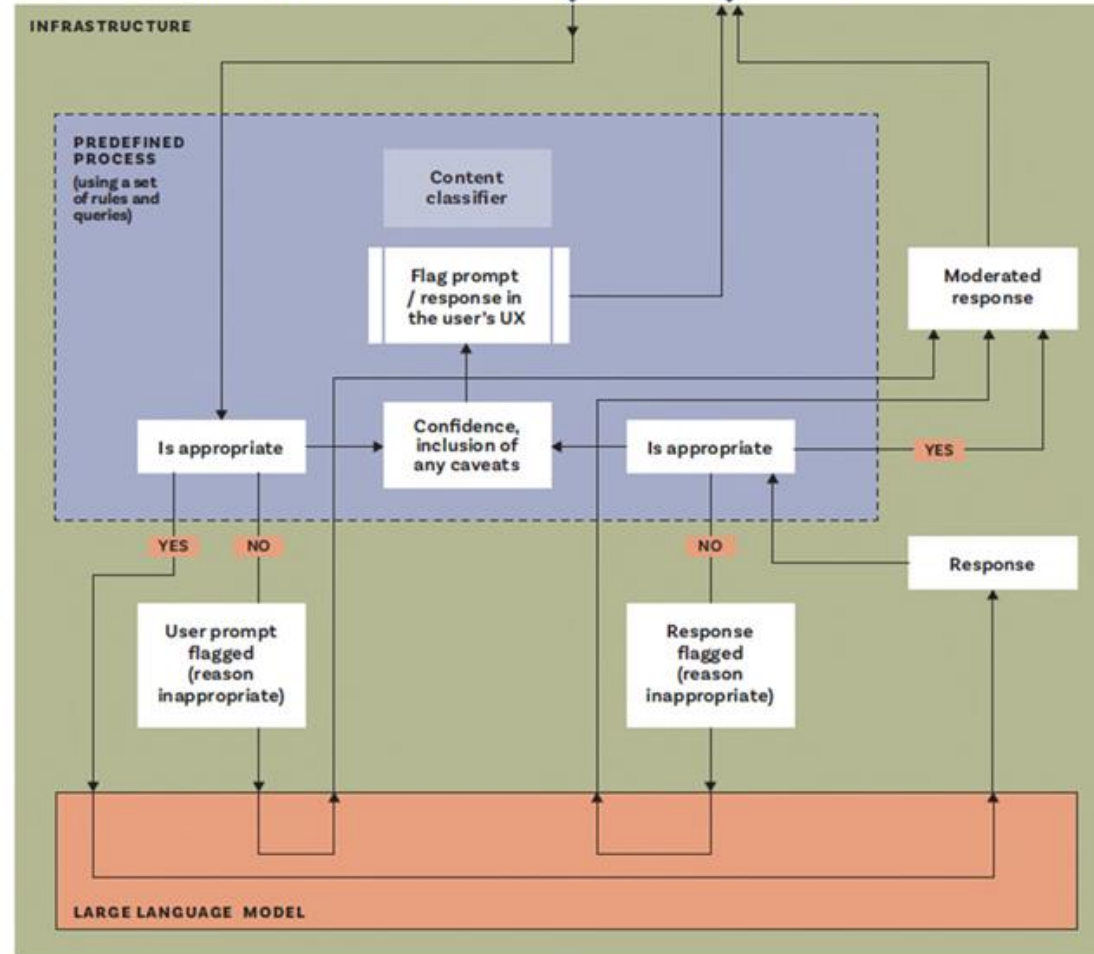
What ChatGPT does

ChatGPT selects its responses from a pre-trained Large Language Model (LLM).
An LLM is an AI designed to understand and generate human-like language.
The current model (GPT3.5) has 175 billion parameters and three billion words.
The ChatGPT application shapes its output based on pre-determined rules and previous interactions.

Typical ChatGPT user experience



Chat GPT3.5 operating process



An example: Robodebt

Within six months of the Robodebt scheme (the Scheme) being launched, it was being heralded as a technological triumph. The Hon Alan Tudge MP, Minister for Human Services, issued a media release on 23 November 2016 titled *New technology helps raise \$4.5 million in welfare debts a day*. The release praised a “new online system” that “is now initiating 20,000 compliance interventions a week – a jump from 20,000 a year... this is a great example of the Government using technology to strengthen our compliance activities with faster and more effective review systems.”²

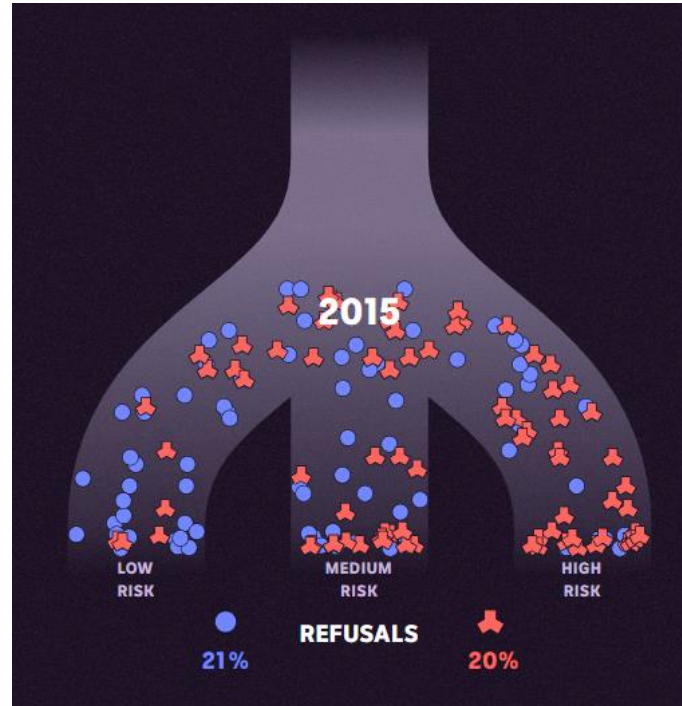
Colleen Taylor, a former employee of DHS, who worked for a period in the Online Compliance team, told the Commission that the first three cases she reviewed when she was employed by that team involved an inadvertent duplication of employer details, so that the same income was counted twice. Ms Taylor said that when her team raised the fact that the debts were incorrect, they were told that their job was to just check that the way the system calculated the debt was correct, not whether the existence of the debt was correct.⁶⁴

The automation used in the Scheme at its outset, removing the human element, was a key factor in the harm it did. The Scheme serves as an example of what can go wrong when adequate care and skill are not employed in the design of a project; where frameworks for design are missing or not followed; where concerns are suppressed,¹⁴⁷ and where the ramifications of the use of the technology are ignored.

An example: the UK Home Office



Risk factors included your skin tone in your photo – darker skin tones were reported to be flagged as having “poor image quality”



With all other variables held constant, the algorithm resulted in applicants from Country B being significantly more likely to have their visa refused



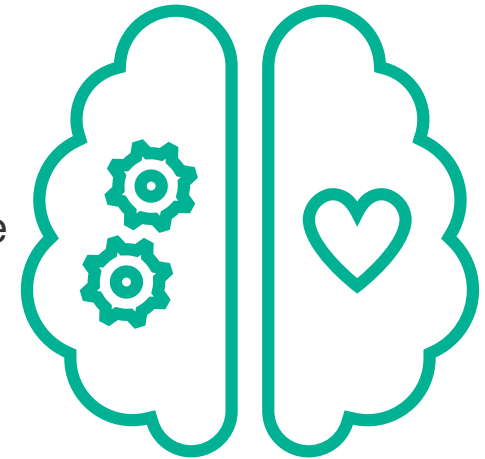
What does this mean for evaluation?

more and more AI and ADM
used in programs

we need to understand how the
technology works not only to
evaluate it, but to be part of the
ethical safeguards surrounding
its proper use

Parameters for evaluating AI-enabled programs

1. evaluate at all stages of the program lifecycle:
 - model design
 - training
 - testing
 - implementation
 - post-implementation
2. evaluate more than usual to understand where the human is in the system, and how oversight, governance and the avoidance of bias is being managed (also with special attention paid to the training data)
3. determine the impact as you would for any other evaluation, but with a particular focus on:
 - the quality of the training and/or input data
 - the decision-making process
 - the impact on the human, considering both positive and negative outcomes
4. but this only works if you understand the technology yourself



Some tips

- › it's always handy to have a sense of better practice, so that you can assess the program under evaluation against better practice features
- › in this context, be aware of and familiar with:
 - Australia's AI Ethics Framework and Principles
 - eSafety Commissioner's Safety by Design approach, including principles and assessment tools
 - DTA and DISR's Interim guidance for agencies on government use of generative Artificial Intelligence platforms
 - IBM Policy Lab Precision Regulation for Artificial Intelligence recommendations
 - OECD AI principles and recommendations
 - EU AI Act
- › a lot to be across?
 - just work on building your knowledge slowly and finding resources that are of use to you
 - webinars are really useful for immersion on the topic. CSIRO's National AI Centre is a great place to start!
 - jargon is jargon, no matter the field – remember how complicated program evaluation is to a layperson? And yet you've learned that, you can learn this too



Free resources to help you



DIY
Program
Evaluation



Build capacity
and culture



Build your own
monitoring
and evaluation
framework



GrosVenorTM
PUBLIC SECTOR ADVISORY

Quality
Thinking.
Quality
Outcomes.

More free resources

Planning a program evaluation

A practical checklist

Grosvenor™
PUBLIC SECTOR ADVISORY



Monitoring and evaluation framework template

Grosvenor™
PUBLIC SECTOR ADVISORY



Grosvenor™
PUBLIC SECTOR ADVISORY

Quality
Thinking.
Quality
Outcomes.

We would love to connect!



Kristy Hornby

Associate Director and Victorian
Program Evaluation Lead

kristyhornby@grosvenor.com.au

03 9616 2700

Grosvenor[™]
PUBLIC SECTOR ADVISORY

Quality
Thinking.
Quality
Outcomes.

For further information
please visit our website:
www.grosvenor.com.au

Contact Us
(02) 6274 9200

 @Grosvenor Public Sector Advisory

Grosvenor[™]
PUBLIC SECTOR ADVISORY

Quality
Thinking.
Quality
Outcomes.